Correcting for primary study misspecifications in meta-analysis

Koetse, M.J.; Florax, R.J.G.M.; de Groot, H.L.F.

2005

document version
Early version, also known as pre-print

TI 2005-029/3
Tinbergen Institute Discussion Paper

Correcting for Primary Study Misspecifications in Meta-Analysis

Mark J. Koetse[a,c]
Raymond J.G.M. Florax[a,b]
Henri L.F. de Groot[a,c]

[a] *Department of Spatial Economics, Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam,*
[b] *Regional Economics Applications Laboratory (REAL), University of Illinois, Urbana-Champaign, USA,*
[c] *Tinbergen Institute.*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Correcting for Primary Study Misspecifications in Meta-Analysis

Mark J. Koetse [a,*], **Raymond J.G.M. Florax** [a,b], **Henri L.F. de Groot** [a]

[a] *Tinbergen Institute and Department of Spatial Economics, Free University, Amsterdam, The Netherlands*

[b] *Regional Economics Applications Laboratory (REAL), University of Illinois, Urbana-Champaign, USA*

Abstract
Misspecifications and differences in operational definitions of elasticities in primary studies carry over to meta-analysis results. We show that the current practice of accounting for such primary study aberrations in a meta-analysis by means of dummy variables goes a long way in mitigating their negative effects on the bias and mean squared error of the estimator, and the size and the power of the statistical tests on the meta-estimate. Controlling for omitted variable bias has a bigger beneficial impact on the meta-analysis results than the concomitant procedure for point versus double-log elasticities. However, the impact of mixing different types of elasticities on the results of a meta-analysis is smaller in any case.

Key words: Meta-analysis; Monte Carlo simulation; Omitted variable bias; Elasticities; Model Misspecification
JEL-codes: C12; C15; C40

---

1. Introduction

Meta-analysis is essentially the 'analysis of analyses' (Hunter and Schmidt, 1990, p. 478). It is a form of research synthesis in which previously documented empirical results are combined or re-analysed in order to increase the power of statistical hypothesis testing. Some proponents maintain that meta-analysis can be viewed as *quantitative* literature review (Stanley, 2001), while others assert that meta-analysis can be used to pinpoint aspects critical to the future development of theory (Goldfarb, 1995; Rosenthal and DiMatteo, 2001). Although meta-analysis was originally developed in experimental medicine, it soon extended to fields such as biomedicine and experimental behavioural sciences. Meta-analysis is currently also gaining ground in economics (see, e.g., Smith and Huang, 1995; Card and Krueger, 1995; Görg and Strobl, 2001; Bateman and Jones, 2003; Weichselbaumer and Winter-Ebmer, 2003).

[*] Corresponding author: Department of Spatial Economics, Free University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Tel.: +31 (0)20 598 6049; fax: +31 (0)20 598 6004. *E-mail:* mkoetse@feweb.vu.nl

The switch from the experimental context to the non-experimental context prevailing in most areas in economics induces specific methodological problems. First, the measurement of effects is less clear-cut. A meta-analysis pertains to the analysis of statistical summary indicators, usually referred to as 'effect sizes.' In experimental sciences, the use of correlations, odds-ratios, and standardised mean differences between experimental and control groups, is customary. Not only are these effect sizes by nature rather different from the typical quantitative measures used in economic research, effect size definitions in economics also tend to be less homogeneous. For instance, elasticities can be measured as a point-elasticity or, alternatively, the true elasticity can be assumed constant across the demand or supply curve and can be directly derived from a double-log specification. Economic meta-analyses typically contain a mix of both types of elasticities (see, e.g., Dalhuisen et al., 2003).

Second, in an experimental research design, sampling of sizeable experimental and control groups mitigates the need for control variables. The design is as a result largely standardised. This is different in economics, where slight changes in research design are often viewed as an innovation over earlier work.[1] Typically, data constraints as well as the desire to be 'different' lead to varying sets of control variables across studies. This obviously induces omitted variable bias in a subset of the set of primary studies in a certain research area. Moreover, in a strict sense, effect sizes estimated with different sets of control variables cannot be assumed to represent identical population effect sizes. However, in a meta-analysis the differences in model specifications are observable across studies, and hence can be controlled for in some way.

Heterogeneity in operational definitions of effect sizes as well as varying sets of control variables are at best taken into account by using appropriately defined dummy variables in the meta-regression specification. This assumes that the impact of omitted variables and of different effect size definitions on the meta-results are linear and additive which, most likely, they are not (see Smith and Pattanayak, 2002). In this paper, we investigate the impact of both types of primary study aberrations on meta-analysis using a Monte Carlo setting.

In the next section we analyse the impact of omitted variables on primary study results and the possible consequences for meta-analysis. Concerning the latter, important criticism has come to the fore about pooling effect sizes from different primary model specifications. In Section 3 we discuss the experimental design of our analysis. Section 4 presents bivariate simulation results in order to illustrate the isolated impact of primary study misspecifications. Section 5 discusses the response surfaces, in which full variation on all variables is induced. Section 6 rounds off with conclusions and ramifications.

---

[1] In medicine, or more specifically epidemiology, the same tendency is apparent in so-called observational studies, which are generally considered less homogeneous than randomised clinical trials (see Sutton et al., 2000).

2. DIFFERENCES IN PRIMARY MODEL SPECIFICATIONS

2.1 Pooling estimates from differently sized models

A rather fundamental problem in meta-analysis associated with differences in primary model specifications is addressed by Keef and Roberts (2004). They argue that effect sizes from models with a different number of explanatory variables are essentially incomparable and should not be pooled in a meta-analysis. Their arguments are centred around the measurement of effect sizes that are common in medicine and psychology, and although they are valid in this specific academic area, they need not hold in the field of economics. Their argument runs through a typical primary model in psychology, which is usually concerned with measuring the mean difference between an experimental and a control group on a certain variable. This model may look as (see also equation 3 in Keef and Roberts, 2004)

$$L = \lambda_0 + \gamma E + \sum_i \lambda_i r_i + \mu \,, \tag{1}$$

where L is the dependent variable. In medicine this could be the mean difference in the number of cured people between an experimental group that was administered a certain medicine and a control group that was not administered this medicine. Similarly, in psychology L could be the mean difference in alcohol addiction between an experimental group that was given a certain treatment and a control group that did not receive this treatment. Furthermore, λ_0 is a constant and γ is the mean difference on L between the experimental and the control group, with E a dummy variable equal to one for the experimental group. Finally, λ_i are coefficients on exogenous explanatory variables r_i, with i = 1,…,k, and μ is an error term. The effect size T resulting from this model is given by

$$T = \frac{\hat{\gamma}}{\hat{\sigma}_\mu^2} \,, \tag{2}$$

where $\hat{\gamma}$ is an estimate of γ and $\hat{\sigma}_\mu^2$ is the estimated variance of μ. The latter is a proxy for the unknown population variance and its purpose is to make T invariant to scale and hence comparable across studies. Suppose a meta-analysis is done on T_j effect sizes that are got from two distinct groups of primary studies. Group one uses a model with number of explanatory variables $k = q$, and group two uses a model with number of explanatory variables $k = r$, with $q < r$. The commonly used meta-model looks as

$$T_j = \theta_0 + \theta_1 D_j + \varsigma_j \,, \tag{3}$$

where θ_0 is the mean effect size of group one, θ_1 is the difference between θ_0 and the mean effect size of group two with D_j a dummy variable equal to one when a study belongs to group two, and ς_j

is an error term. The central point in the Keef and Roberts paper is centred around the fact that $\hat{\sigma}_\mu^2$ in (2) decreases invariantly as the number of explanatory variables in the primary model increases from q to r. The qualitative interpretation of results from the meta-analysis in (3) is still straightforward when $\hat{\theta}_1 < 0$, since the inclusion of $(r - q)$ extra explanatory variables unambiguously leads to a decrease in the mean effect size – despite the decrease in $\hat{\sigma}_\mu^2$. However, when $\hat{\theta}_1 > 0$, unambiguous interpretation is impossible since the decrease in the mean effect size may be due to the impact of the additional explanatory variables or due to the systematic decrease in $\hat{\sigma}_\mu^2$.

Although this issue poses a substantial problem in psychology, it is only partly relevant for meta-analysis in economics. Since scaling in economics occurs by taking elasticities, $\hat{\sigma}_\mu^2$ is not included in economic effect sizes. This basically eliminates the ambiguity in interpretation when $\hat{\theta}_1 > 0$. Potentially problematic at first sight is the fact that $\hat{\sigma}_\mu^2$ is included in the estimation of the standard errors of primary study effect sizes.[2] However, we argue that a decrease in $\hat{\sigma}_\mu^2$ when models get larger is actually a desirable property in meta-models that include standard errors in their estimation. There are two reasons for this.

First, a primary study that does not include all relevant explanatory variables generally produces a less reliable effect size than a study that does, ceteris paribus. When the DGP of primary data is known, a strategy in meta-analysis could be to exclude those studies. However, the DGP of primary data is usually not known in practice. The best strategy in this case is to include all studies and give effect sizes from the studies with omitted variables a weight in the estimation of the meta-model that is proportional to its estimated precision. Because the primary model with omitted variables has less explanatory power and therefore higher residual variance, $\hat{\sigma}_\mu^2$ is higher in such studies. As shown in Sutton et al. (2000), optimal weights in weighted meta-models are defined by $1/\hat{\sigma}_\mu$, i.e., the inverse of the standard error of the primary model estimates, implying that studies with omitted variables indeed get a lower weight in these estimations.

Second, an overspecified primary model produces equally reliable effect sizes as a model that includes the relevant explanatory variables only, ceteris paribus. In this case both studies should get identical weights, which poses no problem since the decrease of $\hat{\sigma}_\mu^2$ in the overspecified model is small, because the extra variables included in this model have, by definition, no explanatory power. Moreover, the small decrease in $\hat{\sigma}_\mu^2$ is fully accounted for, because the sum of squared residuals is divided by degrees of freedom. The latter is of course smaller in the overspecified model ceteris paribus, thereby slightly inflating $\hat{\sigma}_\mu^2$ and giving the effect size that is estimated from this model a slightly smaller weight.

---

[2] Although standard errors are not used in estimation in this paper, they are crucial elements in more sophisticated meta-estimators. These models are introduced in subsequent research.

2.2 Impact of omitted variables on primary study results

Although the pooling of effect sizes from differently sized primary models should pose no fundamental problem in economic applications, this is not to say that differences in the specification of primary models do not have an impact on meta-analysis. Especially when exclusion of relevant explanatory variables lead to bias in primary effect sizes, the results of a meta-analysis may be influenced in several ways. It is precisely these effects that we want to investigate. The easiest way to demonstrate the potential impact of omitted variables on primary study results is by constructing a true underlying model with two explanatory variables. This model may look as

$$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon\,, \tag{4}$$

where Y is the dependent variable, β_1 and β_2 are coefficients on exogenous explanatory variables X_1 and X_2, respectively, and ε is an error term. When this model is estimated by OLS, producing $\hat{\beta}_1$ as an estimate of β_1, then under well known conditions we have $E(\hat{\beta}_1) = \beta_1$. Our interest is in the situation where the model in (4) is estimated without X_2 as an explanatory variable. In this case we get $\hat{\beta}_1^{ov}$ as an estimate of β_1, which has an expected value equal to (see Greene, 2000, p. 335)

$$E\left(\hat{\beta}_1^{ov}\right) = \beta_1 + \beta_2 \frac{Cov\left(X_1, X_2\right)}{Var\left(X_1\right)}\,, \tag{5}$$

which clearly shows that $\hat{\beta}_1^{ov}$ is biased if $Cov(X_1, X_2) \neq 0$. Although the sign of the bias in this model is equal to the sign of the covariance times the sign of β_2, the sign of the unknown covariance is often unclear in reality, especially in aggregate data. Therefore, even if β_2 is known, we are left with a bias in $\hat{\beta}_1^{ov}$ of unknown sign and size. As the model in (4) is expanded to include more regressors and is again estimated without X_2, the situation gets even worse. In this case the extended version of (5) involves multiple coefficients that themselves are partial, not simple correlations (Greene, 2000, p. 336), and more covariance terms of unknown sign and size.

The consequences for meta-analysis are potentially serious, and depend primarily on the characteristics of the underlying set of primary studies. An ideal situation would be to have a set of studies with identical model specifications, with similar measures for the endogenous and exogenous variables, and with similar disaggregate data. Not surprisingly, such circumstances are often found only in experimental research areas. Many areas of economic research do not show such an ideal picture however; model specifications are different, measures diverge and data range from micro-economic panel data to aggregate cross-sections. In reality we can at best hope that the impact of omitted variables across primary studies is randomly distributed around a certain mean. If, for instance, $Cov(X_1, X_2)$ is normally distributed around zero across studies, then asymptotically the bias of the primary study effect sizes is zero as well, while the mean squared error is substantially larger than zero. Because our principal aim is to test the simple correction mechanism used in meta-analysis, the experimental design set out in the next section ignores the specific methodological problem set out above. Note, however, that is important and will be analysed in subsequent research.

3. EXPERIMENTAL DESIGN

The set-up of our experimental simulations comprises three steps: (I) generating the primary data, (II) estimating two different versions of the primary model, and (III) performing the meta-analyses using the estimated effect sizes of the primary studies as inputs. Subsequently, we investigate the results of the various meta-analyses in a so-called response-surface analysis in order to draw general conclusions about the bias, the mean squared error and the statistical tests of the meta-estimate of the population effect size.[3] In order to avoid confusion with respect to notation, observe that we redefine the meaning of the variables and parameters used in the previous section. Going short, we start from scratch.

3.1 Primary data and primary models

The 'true' model used for *data generation* of the primary studies is a general unrestricted Cobb-Douglas function of the form

$$y = e^{\alpha} x^{\beta_0} z^{\beta_1} e^{\varepsilon}, \tag{6}$$

where y is a stochastic variate, x and z are exogenous variables, α, β_0 and β_1 are population parameters, and ε is an error term. In the Cobb-Douglas model the elasticity of y on x equals β_0 over the entire data range, and we set β_0 equal to 0, 0.5, 1.0 and 1.5, mimicking the situation where the population effect is zero, and situations in which there is a positive effect, respectively. We fix the parameters α and β_1 at unity, and draw the exogenous variable x once from a uniform (1,10) distribution. In order to ensure that x and z are correlated, we generate z for each replication according to

$$z = x + e^{\psi}, \tag{7}$$

where ψ is a vector containing normally distributed errors with mean zero and variance two. We draw ψ from a random normal distribution with mean and variance such that the correlation between x and z is approximately 0.4. The error term ε in (6) is normally distributed with mean zero, and variance 0.5, 1.0 and 2.0, respectively. Alternatively, we use a log-normal distribution with adequately resized

---

[3] The simulation program used for the analyses in this paper are written in the Gauss statistical software package. The program is available upon request from the authors.

variances for the error term.[4] As primary studies vary in size, we use sample sizes of 50, 100, 500 and 1,000, respectively.

Our approach is different from other Monte Carlo studies in meta-analysis (e.g., Oswald and Johnson, 1998; Bijmolt and Pieters, 2001) because we explicitly incorporate the stage of the *primary data analysis*. First, this enables us to investigate the impact of different operationalisations of the effect size definition on the meta-analysis results. Specifically, we use the data generated by (6) to estimate the log-linear model that is mathematically equivalent to the multiplicative model

$$\ln(y) = \alpha + \beta_0 \ln(x) + \beta_1 \ln(z) + \varepsilon . \tag{8}$$

We estimate this model by OLS, which gives $\hat{\alpha}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ as estimates of α, β_0 and β_1, respectively. The parameter of interest is the double-log elasticity of primary study *k*, given by $\hat{\eta}_k = \hat{\beta}_{0k}$. In this case the elasticity can be viewed as the 'true' elasticity. Note that it is constant across the entire primary data-set since by construction (see equation (6)). The standard error of the elasticity is simply the standard error of $\hat{\beta}_0$ from (8). Alternatively, given *x*, *y* and *z* generated by (6), we use OLS to estimate an additive primary model

$$y = \alpha' + \beta_0' x + \beta_1' z + \varepsilon' \ , \tag{9}$$

producing $\hat{\alpha}'$, $\hat{\beta}_0'$ and $\hat{\beta}_1'$ as estimates of α', β_0' and β_1', respectively. In this model we linearly estimate the non-linear relationship between *y* and *x* and *z*, and compute a point-elasticity at the sample mean, for say primary study *m*, as $\hat{\eta}_m = \hat{\beta}_{0m}' (\bar{x} / \bar{y}_m)$. In reality this may occur frequently, simply because the true underlying model is unknown and researchers may erroneously the true underlying model to be linear. The argument for using the mean values is that most primary studies that calculate point estimates of an elasticity do this at the sample mean, a possible alternative being the median. The standard error of this elasticity is obviously not the standard error of $\hat{\beta}_0'$ from (9). For calculating the standard error we use the Delta method (see Greene, 2000, pp. 359-360). In this case the method dic-

---

[4] Since the mean and variance of the two disturbances have to be identical, to create a lognormal disturbance we take the exponent of a normally distributed error with variance equal to

$\sigma_n^2 = \ln(0.5 + 0.5\sqrt{1 + 4\sigma_l^2})$,

where σ_n^2 is the variance of the normally distributed disturbance and σ_l^2 is the preferred variance of the lognormal disturbance. After this transformation we get

$\mu_l = \exp(0.5 \ln(0.5 + 0.5\sqrt{1 + 4\sigma_l^2}))$,

with μ_l the mean of the lognormal disturbance, so we simply subtract the difference between this number and the preferred μ_l from the attained disturbance term to get the preferred μ_l .

tates that we multiply the standard error of $\hat{\beta}_0'$ from (9) by the same ratio of means as before. This means that $SE(\hat{\eta}_m) = SE(\hat{\beta}_{0m}')(\bar{x}/\bar{y}_m)$ for primary study *m*.

It is not difficult to see that when the underlying set of $\{ y , x , z \}$ combinations is jointly normally distributed, the point-estimate should be a good approximation of the double-log estimate. However, when normality does not hold, the point-estimate and double-log estimate may diverge. The see this note that the *mean elasticity* is likely to be different from the *elasticity at the mean*, and that therefore the mean in a non-normal distribution may not be the best measure of centrality. Since the differences are not tractable analytically, we have to rely on simulations to analyse whether using point-elasticities in meta-analysis actually matters for its results.

The second central issue in this paper is the existence of omitted variable bias in primary studies and its impact on the results of a meta-analysis. Therefore, in order to induce omitted variable bias we estimate the models in (8) and (9) with and without z included among the explanatory variables. This should lead to a bias in primary effect sizes since x and z are correlated by construction.

Summing up, the variations that we induce include four values for β_0, four different sample sizes in primary studies, three values for the disturbance variance, a normal and a log-normal disturbance term, a point- and a double-log elasticity, and presence or absence of omitted variable bias. We therefore end up with 4×4×3×2×2×2 = 384 possible combinations. For each of the possible combinations we run 100 replications, resulting in 38,400 'primary studies' to sample from for the *meta-analyses*.

3.2 Sampling studies for the meta-analyses: A stratified joint probability procedure

The study retrieval process, determining which 'studies' end up in the meta-sample, is based on a stratified two-stage random sampling procedure. Note that simple random draws from the pool of studies, eventually stratified according to the true value of the population effect size, would asymptotically result in meta-analyses with study characteristics on average 'fixed' in the same proportions with which they were generated in the experimental design. The more complicated random sampling procedure described below precludes the experimental set-up driving the simulation results of the meta-analyses, and safeguards that sufficient variation in characteristics of primary studies that are included in the meta-analysis is available. This closely resembles what actually occurs in the practice of doing meta-analysis research.[5]

First, we define four strata according to the true value of the population effect size β_0, since if we put studies with a different value for β_0 in a single meta-analysis, the true value of our meta-

---

[5] Note that an interesting possible modification of the sampling scheme is to account for publication bias (see Stanley, 2001, and Florax, 2002, for an overview).

estimate is unknown. Consequently, we cannot sensibly measure and compare the performances of different meta-models.

Second, for each meta-analysis in a specific stratum we compute a joint sampling probability for each of the 96 study *types* (= 384/4 possible combinations of study characteristics) by randomly drawing sampling probabilities for each study *characteristic* from a uniform (0,1) distribution. With sufficient replications this sampling procedure ensures a maximum variation in sampling probabilities. As an example, suppose we have 6 different study types from 2 study characteristics – omitted variable bias (OVB, 2 alternatives) and disturbance variance (DV, 3 alternatives). We draw the probability for sampling a study with omitted variables bias from a uniform (0,1) distribution, for example 0.35. The probability of sampling a study without omitted variable bias is automatically (1–0.35)=0.65. Furthermore, the probability of sampling a study with a certain disturbance variance is determined as follows. We draw for each disturbance variance from a uniform (0,1) distribution, for example 0.25, 0.60 and 0.40. Since the sum of probabilities is not equal to 1, we divide each probability by the sum of the three probabilities, resulting in three sampling probabilities that do sum to 1, in this case 0.20, 0.48 and 0.32.

Third, we multiply the joint sampling probability of each specific study type with the sample size of the meta-analysis, resulting in the absolute number of replications of each study type to be included in the meta-analysis. We then accordingly sample the studies to be included in the meta-analysis randomly from the available replications per study type. The joint sampling probability and the absolute number of studies for each of the 6 different study types is given in Table 4.1. In this example the sample of the meta-analysis consists of 100 primary studies. Note that the joint sampling probabilities and number of studies sum to 1 and 100 respectively.

Table 4.1. Example of the sampling procedure; joint sampling probabilities for 6 different study types

Study type	Probability OVB	Probability DV	Joint probability	Number of studies
OVB – DV 1	0.35	0.20	0.07	7
OVB – DV 2	0.35	0.48	0.17	17
OVB – DV 3	0.35	0.32	0.11	11
No OVB – DV 1	0.65	0.20	0.13	13
No OVB – DV 2	0.65	0.48	0.31	31
No OVB – DV 3	0.65	0.32	0.21	21

3.3 Meta-models and -estimations

We allow the sample size of the meta-analysis to take on values of 25, 50, 100 and 200, and in order to attain sufficient accuracy we run 5,000 meta-analyses. The latter implies that we draw 5,000 times from the pool of primary studies using the sampling procedure set out above, after which we do a meta-analysis for each of these 5,000 meta-samples. The number of meta-results to analyse, resulting from the use of four different values for the true underlying effect β_0, four different meta-sample sizes, and 5,000 meta-analyses for each combination, is equal to 80,000. We estimate two meta-models. The first model is a rather naïve meta-regression specification in which we take the mean of the estimated effect sizes as an estimator of the population effect size:

$$\hat{\eta}_s = \delta_0 + \xi_s \,, \tag{10}$$

where $\hat{\eta}_s$ is a vector with a mix of double-log and point-elasticities, δ_0 is a constant with expected value equal to the population effect size β_0 of the multiplicative model if the estimator is unbiased, ξ_s is a residual term, and s = 1,…,S is the number of primary studies. This model is estimated by OLS giving us $\hat{\delta}_0$ as the estimate of δ_0. Using OLS, $\hat{\delta}_0$ and $Var(\hat{\delta}_0)$ are, respectively, given by:

$$\hat{\delta}_0 = \frac{1}{S}\sum_{s=1}^{S}\hat{\eta}_s \,, \tag{11}$$

and

$$Var\left(\hat{\delta}_0\right) = Var\left(\frac{1}{S}\sum_{s=1}^{S}\hat{\eta}_s\right) = \frac{1}{S^2}\sum_{s=1}^{S}Var\left(\hat{\eta}_s\right) = \frac{1}{S^2}\sum_{s=1}^{S}\sigma^2_{\hat{\eta}_s} \,. \tag{12}$$

The variance in (12) clearly shows one of the advantages of meta-analysis. As S increases, the variance of the meta-estimate decreases. Moreover, the variance of the meta-estimate very quickly becomes substantially lower than the smallest existing variance in primary studies, even for relatively small meta-samples. Furthermore, when $E\left(\hat{\eta}_s\right) = \beta_0 \ \forall\ s$, it follows from (11) that $E(\hat{\delta}_0) = \beta_0$. When $E(\hat{\eta}_s) \neq \beta_0$ for any i, than $E(\hat{\delta}_0)$ may still be equal to β_0, if the differences average out. However, the variance of the estimate is always larger in this case. Alternatively, if differences do not average out, $\hat{\delta}_0$ is clearly biased since $E(\hat{\delta}_0) \neq \beta_0$. In this case, the variance of the estimate may remain unchanged, however. As we have induced two types of misspecification in primary studies, we virtually assured that $E(\hat{\eta}_s) \neq \beta_0$ for some i.

A model that is frequently employed to account for primary study misspecifications is a meta-regression model in which dummy variables control for observable differences between primary study characteristics. This model is given by:

$$\hat{\eta}_s = \delta_0' + \delta_1' D_s^{ov} + \delta_2' D_s^{pe} + \xi_s' \,, \tag{13}$$

where D_s^{ov} is a dummy variable equal to one if the primary study is estimated without *z* among the explanatory variables, and D_s^{pe} a dummy variable equal to one if the effect size of the primary study is a point-elasticity. We also estimate this model by OLS, the relevant parameter in this case being $\hat{\delta}_0'$ as the estimate of δ_0'. One of the most important assumptions underlying this model is that the relationship between differences in primary study characteristics on the one hand, and differences in primary study effect sizes on the other, is linear. If this holds, which is by no means guaranteed, then the changes in effect sizes due to primary model misspecifications should be largely accounted for by the dummy variables. When the relationship is non-linear, the effectiveness of the proposed solution is uncertain. Alternatively, if the impact is stochastic around zero, which in practice is not likely for omitted variable bias but may be the case for point-elasticities, differences may average out, especially in large samples. In this case the impact is not picked up by the dummy variables and the model in (13) may in effect converge to the model in (10).

3.4 Measuring model performance

The pivotal issue at stake is how well the meta-analyses recover the value of the population effect size β_0 in terms of size and statistical significance. Specifically, we compare the 'naïve' meta-regression estimate $\hat{\delta}_0$ from (10) to $\hat{\delta}_0'$, the estimate from the meta-regression specification including dummy variables to control for omitted variable bias and mixing two types of elasticities (see equation (13)).

As discussed previously, the misspecifications may affect the meta-estimates on several dimensions, i.e., the estimate itself, the variance of the estimate, and the significance of the estimate. We therefore use three different performance indicators to investigate the impact. First, we use numerical approximations for the bias (BIAS) of the meta-estimates $\hat{\delta}_0$ and $\hat{\delta}_0'$, i.e., whether the average value of the estimates differs from β_0. As we have argued in the previous subsection, a problem with this indicator is that the impact of misspecifications on the effect sizes may average out, in which case bias is equal to zero. However, the variance of the estimate may still be substantial. In order to account for this we also use the mean squared error (MSE) of the estimate as a performance indicator. This second indicator combines the bias and the variance of the estimate, and measures the average distance of the estimate to the true parameter, i.e., the smaller the MSE, the closer the estimate will be to the true parameter, on average. The third and final indicator is the proportion of statistically significant results (SIG) of the meta-estimators. Formally, for $\hat{\delta}_0$, these indicators are given by:[6]

[6] The performance indicators for $\hat{\delta}_0'$ are obtained by replacing $\hat{\delta}_0$ in (14), (15) and (16) by $\hat{\delta}_0'$.

$$\mathrm{BIAS}\left(\hat{\delta}_0\right) = \mathrm{E}\left(\hat{\delta}_0 - \beta_0\right) \approx \frac{1}{M}\sum_{m=1}^{M}\left(\hat{\delta}_0 - \beta_0\right)_m , \tag{14}$$

$$\mathrm{MSE}\left(\hat{\delta}_0\right) = \mathrm{E}\left(\hat{\delta}_0 - \beta_0\right)^2 = \left[\mathrm{BIAS}\left(\hat{\delta}_0\right)\right]^2 + \mathrm{Var}\left(\hat{\delta}_0\right) \approx \frac{1}{M}\sum_{m=1}^{M}\left(\hat{\delta}_0 - \beta_0\right)_m^2 , \tag{15}$$

$$\mathrm{SIG}\left(\hat{\delta}_0\right) = \frac{1}{M}\sum_{m=1}^{M} I\left(\left|t_{n-k}\right| > \tau\right)_m , \tag{16}$$

where I in (16) is an indicator function that is one if the absolute *t*-value is greater than a pre-specified critical t-value denoted by τ, and 0 otherwise; $m = 1,\ldots,M$ indexes the meta-analyses. In what follows, we apply two-sided significance tests using a 5% significance level. When $\beta_0 = 0$ and $H_0\colon\ \beta_0 = 0$, we are interested in the probability of a Type I error, i.e., the probability that an estimator erroneously rejects H_0. Therefore, SIG corresponds to the proportion of Type I errors or the *size* of the meta-estimator when $\beta_0 = 0$. Alternatively, when $\beta_0 \neq 0$ and under the same null-hypothesis, we want to know the probability of a Type II error, i.e., the probability that an estimator erroneously accepts H_0. Therefore, SIG corresponds to 1 minus the probability of a Type II error, or the *power* of the meta-estimator when $\beta_0 \neq 0$. Since erroneously rejecting the null-hypothesis requires a considerably larger confidence interval than erroneously accepting the null-hypothesis, the two indicators are not reciprocal and provide different types of information on an estimator. However, the two tests are clearly related. For instance, as the size of an estimator decreases, implying relatively small standard errors, the probability of large power of the estimator increases.

4. Impact of Misspecifications on Meta-Analysis: Bivariate Illustrations

Before turning to the response surface of our simulation exercise, in which full variation is induced on all the variables introduced in Section 3, we focus on bivariate simulations in the current section. Specifically, we vary the number of point-elasticities and the number of primary studies with omitted variable bias in the meta-samples, and keep constant all other variables. This procedure isolates the impact of both misspecifications on the results of a meta-analysis. It furthermore allows us to compare very precisely the performance of the two meta-regression models in (10) and (13).

The results are generated for primary studies with sample size 100 and a standard normally distributed error term. The number of replications for the primary studies is 500, the sample size of the meta-analysis is 50 and the number of meta-analysis replications is 40,000. We distinguish between the situation where the true underlying effect size is zero and one, respectively. Figure 4.1 shows the three performance indicators for the naïve meta-regression model specification in (10) for both situations. On the horizontal axis, we measure the extent of misspecification, representing either the proportion of primary studies included in the meta-analysis suffering from omitted variable bias or, alternatively, the proportion of primary studies included in the meta-sample that uses point-elasticities. In each plot, we depict four ‘extreme’ cases:[7]

- Increasing proportion of studies with omitted variable bias with 5% of the primary studies containing point-elasticities (black square);
- Increasing proportion of studies with omitted variable bias with 95% of the primary studies containing point-elasticities (white square);
- Increasing proportions of point-elasticities with 5% of the primary studies containing omitted variable bias (black triangle);
- Increasing proportions of point-elasticities with 95% of the primary studies containing omitted variable bias (white triangle).

The proportion of misspecification in a meta-sample is given in deciles. With respect to the meta-samples with an increasing number of primary studies containing omitted variable bias, results for the fourth decile, for instance, refer to the BIAS, MSE and SIG for meta-analyses in which the number of studies with misspecification is between 30% and 40%. Furthermore, note that on each indicator the black triangle and black square coincide at the first decile, while their white counterparts coincide at the tenth decile. This result is not surprising since at these deciles the characteristics of the meta-samples are almost identical on average.

With respect to the bias of the estimator in Figure 4.1, mixing different types of elasticities (point versus double-log elasticities) induces a relatively small positive bias when the population effect size is zero, as shown by the small difference between the lines with black and white squares, respectively. Furthermore, the shallow slope of both lines implies that the bias varies only slightly according to the proportion of point-elasticities included in the meta-sample. Omitted variables in primary studies appear much more of a problem, judging by the difference between the lines with white and black triangles, respectively. Its impact is similar when the population effect is one. Under these circumstances the bias induced by mixing different type elasticities is also much more severe. Moreover, it is now strongly negatively correlated with the proportion of point-elasticities included in the sample.

The results for the mean squared error are similar, and show that there is usually no apparent trade-off between the bias and variance of the meta-estimator (see also Greene, 2000, p. 104). At first sight it appears that an exception pertains to the situation in which the primary studies suffer from

[7] In these extreme cases we use 5% and 95% because using 0% and 100% would lead to models in which one of the dummy variables would have to be excluded because they would have no variation or would be perfectly collinear with the constant.

omitted variable bias and exclusively use point-elasticities; an increasing bias seems to be partly offset by a lower variance of the meta-estimator, while the reverse seems to be true for meta-analyses with more than 50% primary studies with omitted variable bias. Note, however, that the bias of the estimate is negative and increasing up to the fifth decile, and therefore not increasing in absolute terms. By simply taking the square of the bias we get the observed U-shaped MSE curve, implying that there is no trade-off after all.

The results for SIG show that the power of the naïve meta-estimator is satisfactory, whereas the size of the estimator is seriously affected by the misspecifications. Specifically, omitted variables have a strong effect, but even mixing different type elasticities (without omitted variable bias) results in a size that is too large.

Figure 4.2 presents the results for the meta-regression model with the dummy variable correction procedure given in (13). We focus our discussion on the comparison of the two meta-models. It is evident that the dummy correction procedure substantially reduces the bias and mean squared error of the meta-estimator for both population effect sizes. The BIAS and MSE are virtually zero for the case in which the misspecifications occur in isolation (either omitted variable bias or inclusion of point-elasticities in the meta-sample). When both misspecifications are present there is an interaction effect that is somewhat more substantive, although the bias and mean squared error are still small in absolute terms and relative to the 'naïve' estimator.

Finally, the size of meta-estimator using the dummy correction procedure is much more in accordance with what it should be, in our case around the 0.05 level. In three of the four cases the size is acceptable, unless the misspecifications are particularly strong (deciles 9 and 10). When all studies suffer from omitted variable bias the size is well above the 0.05 level. Still it is far more acceptable than the size of the 'naïve' estimator in this situation, which was constant at 1. The power of the meta-estimator with dummy corrections is again satisfactory, although in extreme cases it is not as good as the power of the 'naïve' estimator. In fact, the power of the estimators is the only indicator on which the meta-model with dummy correction performs worse on some accounts than the 'naïve' meta-model.

5. RESPONSE SURFACES

We proceed by analysing the BIAS, MSE and SIG as the dependent variables in a response-surface analysis in order to arrive at more generally valid conclusions about the small sample behaviour of the two meta-estimators. We estimate separate response surfaces for the 'naïve' meta-estimator and the meta-estimator with dummies to correct for the two types of primary model misspecifications. We distinguish between the situation where the population effect size is zero and where it is larger than zero (taking on values of 0.5, 1.0 and 1.5). The set of exogenous variables contains dummy variables for the magnitude of the population effect size (except in the case where $\beta_0 = 0$), the size of the meta-sample, and the log of the average sample size of primary studies in the meta-sample. We also utilise variables representing the percentage of primary studies included in the meta-sample with point-elasticities, omitted variable bias, a log-normal error distribution, and a specific error variance, respectively.

Graphical depictions of the dependent variables (BIAS, MSE and SIG) suggest that the error variance of the response surface regression is heteroskedastic. We furthermore suspect that some clusters of covariances are interdependent, simply because the data used for the meta-analyses within these clusters are similar. Both issues render OLS estimation inefficient. We therefore define clusters according to the main dimensions of variation. Specifically, we use the value of the population effect size, the meta-sample size, and the number of point-elasticities in a meta-analysis (measured in deciles) to determine the clusters (see also the corresponding note to Table 4.2). We estimate the response surfaces for BIAS and MSE using the clustered Huber-White sandwich estimator, thereby simultaneously correcting for heteroskedasticity and within-cluster dependency (see Williams, 2000; Wooldridge, 2002, Section 13.8.2). The analysis on SIG is based on binary probit estimation utilising a similar robustness device. Note that, with respect to the size of the test (SIG when $\beta_0 = 0$), the coefficient on 'Sample size meta-analysis = 200' does not converge because it contains almost no variation, i.e., the power for large sample meta-analyses approaches 1. We estimate the model excluding this variable as a regressor.

Table 4.2 presents the results for the response surfaces. For the naïve meta-estimator, increasing the number of point-elasticities in a meta-sample has a negative effect on both BIAS and MSE. The impact increases considerably for population effect sizes larger than zero. Even more serious is the effect of increasing the number of studies with omitted variables in a meta-sample, which has a strong positive effect on both indicators. With respect to the analysis on SIG, a negative coefficient implies a decrease in the probability of a Type I error (increase in size) when $\beta_0 = 0$, and an increase in the probability of a Type II error (decrease in power) when $\beta_0 > 0$. Thus, point-elasticities have a positive impact on the size but a negative impact on the power of the test. For studies with omitted variables this is exactly the other way around. Although the latter was already clear from the previous section, the former was not.

Once the meta-estimator with dummy variables is used, the effect of increasing proportions of point-elasticities on BIAS and MSE becomes positive. In effect, the use of dummy variables overcorrects the negative BIAS and MSE of the naïve meta-estimator. The effect of omitted variable bias in terms of BIAS and MSE of the meta-estimator using dummy variables is substantially reduced, and the effects of the two misspecifications are comparable in magnitude. However, as already suggested by results in the previous section, the meta-estimator using dummy variables does not fully mitigate the effect of the two misspecifications. Furthermore, the results on the SIG indicator again show that the use of the estimator with dummy variables is not advantageous on all accounts. Although the negative impact of omitted variables on the size of the test is reduced considerably, its positive effect on the power of the test is reversed. Moreover, the positive impact of point-elasticities on the size of the test has decreased, while its negative impact on the power has increased substantially.

Several other results are interesting as well. First, as expected the BIAS and MSE are larger in absolute value the larger the population effect size, although, when the estimator with dummy variables is used, the effects are very small if not statistically insignificant. Second, the sample size of the meta-analysis does not seem to reduce the bias of the meta-estimators, and causes only a small reduction in the MSE in some cases. For the 'naïve' estimator, increasing the meta-sample substantially increases the probability of a type I error or the size of the test. This is due to the fact that increasing the meta-sample decreases the standard error of the meta-estimate, leading to smaller confidence intervals for any prespecified significance level. Therefore, taking into account that the mean squared error of the meta-estimate is not or only marginally affected by an increase in the meta-sample, the probability that the null hypothesis H_0: $\beta_0 = 0$ is rejected becomes larger as the meta-sample is increased. Although the estimator with dummy variables substantially reduces this negative impact, the effect is still there. By the same reasoning, the power of the estimator increases as the meta-sample is increased for both meta-estimators, which is according to expectation. Third, a larger average sample size of primary studies in the meta-sample decreases the BIAS and MSE of the meta-estimator, although the effect is very small. Increasing average primary study sample size is furthermore beneficial for the size of a test – the effect being stronger for the estimator with dummy variables. It has no impact on the power of the test for both estimators. Fourth, a log-normal distribution has a negative effect if the naïve meta-estimator is used, whereas it is positive for the corrected meta-estimator (except for SIG). Finally, the results for the disturbance variance are slightly mixed, but they generally indicate that for the corrected meta-estimator, sampling primary studies with a better fit (i.e., smaller error variance) lowers the BIAS and MSE. However, the effect is rather small.

6. Conclusions and Ramifications

This paper investigates the impact of common primary study misspecifications on the results of a meta-analysis, and the performance of standard correction procedures in correcting for these aberrations. Although some inferences on this issue can be drawn analytically, we had to rely on simulations for answering most of our research questions.

Our simulations mimic the actual practice in meta-analysis rather accurately. We induce variation in primary studies by including two types of misspecification and estimating the primary models with different sample sizes, error distributions and error variances. Subsequently, primary studies are selected into the meta-sample by using a stratified random probability procedure, ensuring maximum variation given the fact that we use sufficient replications. We distinguish between two meta-models, both to be estimated by OLS; a 'naïve' estimator, for which the meta-estimate is simply the average value of the effect sizes, and an estimator that uses dummy variables to correct for observable primary study misspecifications.

We use three indicators for measuring model performance, i.e., the bias, the mean squared error and the size and power of the meta-estimators. The general conclusion is that, although the effects of misspecifications on the results of a 'naïve' meta-analysis are substantial, common procedures to correct for these aberrations go a long way in reducing the effects on the bias and mean squared error of the meta-estimate. Admittedly, results on the power of the test show that these procedures are not beneficial on all accounts, although the negative effects are small and bivariate results suggest that this is mainly the case for meta-samples with very high proportions of primary studies with misspecification. Moreover, the size of the test for the 'naïve' estimator is unacceptable on all accounts, whereas it is acceptable for the estimator with dummy variables, except for meta-samples with very high proportions of studies with misspecification. Therefore, results from using the 'naïve' estimator simply do not answer the question whether the true underlying effect is different from zero or not, since a significant result is often obtained in any case.

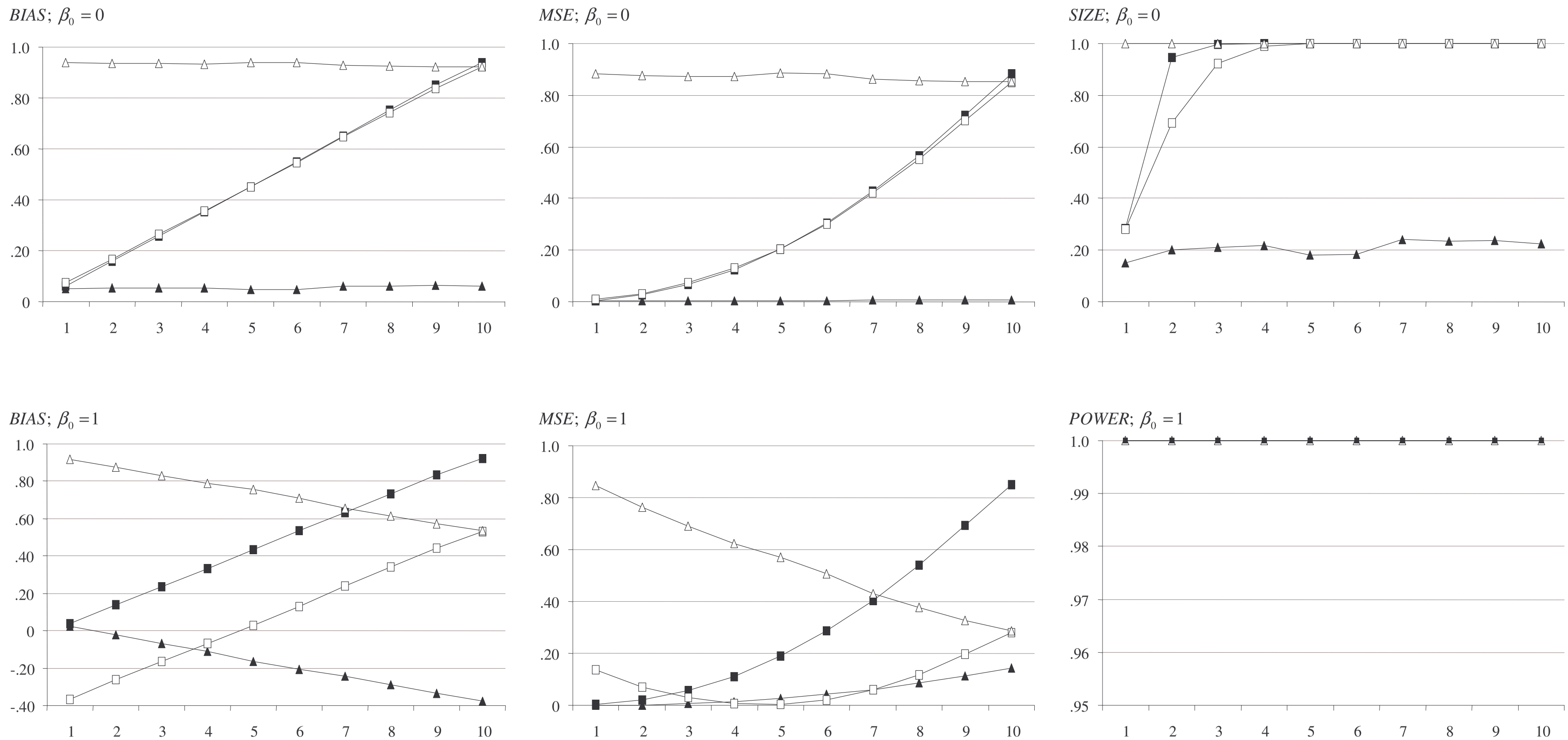

Figure 4.1. BIAS (left), MSE (middle), and SIG (right) in the ‘naïve’ meta-regression specification, for the case where the population effect size $\beta_0 = 0$ (top), and the population effect size $\beta_0 = 1$ (bottom), against the proportion of primary studies included in the meta-sample (in deciles) with omitted variable bias or point-elasticities.

Note: The results are generated for primary studies with sample size 100 and a standard normally distributed error term. The meta-analysis sample size is 50. The number of replications for the primary studies is 500, and the number of meta-analyses is 40,000. The different lines pertain to meta-analyses with increasing omitted variable bias and 5% point-elasticities (black square), increasing omitted variable bias and 95% point-elasticities (white square), increasing proportions of point-elasticities and 5% studies with omitted variable bias (black triangle), and increasing proportions of point-elasticities with 95% with omitted variable bias (white triangle).

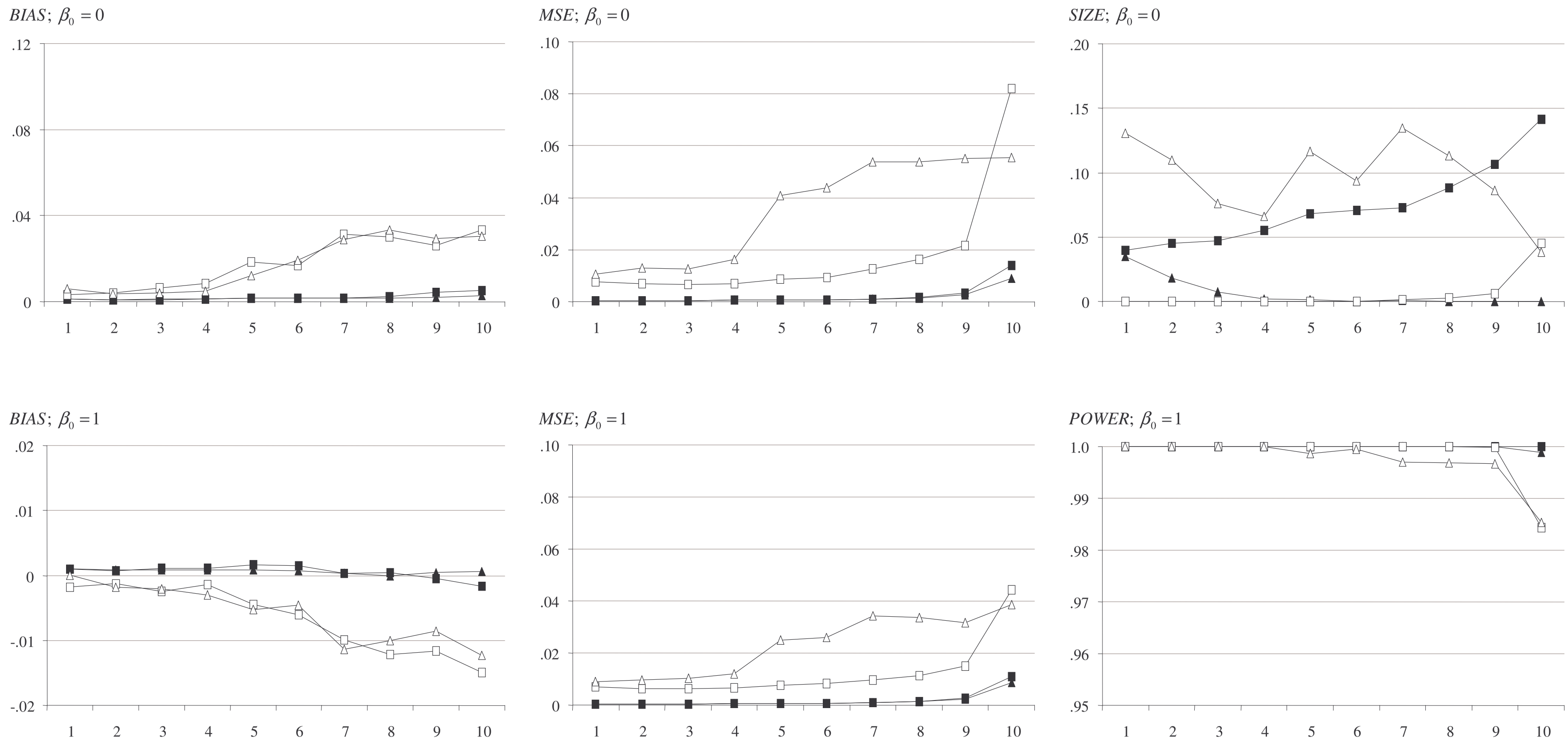

Figure 4.2. BIAS (left), MSE (middle), and SIG (right) in the meta-regression specification with dummy variables, for the case where the population effect size $\beta_0 = 0$ (top), and the population effect size $\beta_0 = 1$ (bottom), against the proportion of primary studies included in the meta-sample (in deciles) with omitted variable bias or point-elasticities.

Note: See the note to Figure 4.1.

Table 4.2. Response surface regressions on BIAS, MSE and SIG for the 'naïve' meta-estimator and the meta-estimator with dummy variables, for $\beta_0 = 0$ and $\beta_0 \neq 0$[a]

Estimator	Naïve meta-estimator						Meta-estimator with dummy variables					
Population effect size	$\beta_0 = 0$			$\beta_0 \neq 0$			$\beta_0 = 0$			$\beta_0 \neq 0$		
Exogenous ↓ \ Dependent variable →	BIAS	MSE	SIZE	BIAS	MSE	POWER	BIAS	MSE	SIZE	BIAS	MSE	POWER
Constant[b]	0.25	0.11	-0.28†	0.02†	0.10	5.55	0.03†	-0.05†	-0.48†	-0.11	-0.07	5.72
	(19.9)	(6.45)	(-1.08)	(0.97)	(5.28)	(9.10)	(1.12)	(-1.39)	(-1.26)	(-7.13)	(-4.13)	(20.2)
$\beta_0 = 1$ (dummy)	--	--	--	0.30	0.10	-1.43	--	--	--	0.02	0.001†	-1.95
				(15.9)	(10.5)	(-5.95)				(6.42)	(0.17)	(-14.5)
$\beta_0 = 1.5$ (dummy)	--	--	--	0.15	0.02†	-0.67	--	--	--	0.02	0.005†	-0.75
				(11.3)	(1.66)	(-3.04)				(6.86)	(0.89)	(-4.98)
Sample size meta-analysis = 50 (dummy)	-0.003†	-0.01	0.98	0.001†	-0.004†	0.67	-0.0004†	-0.03	0.13†	-0.0004†	-0.03	0.75
	(-1.90)	(-3.08)	(15.9)	(0.08)	(-0.43)	(4.04)	(-0.06)	(-2.48)	(1.34)	(-0.10)	(-4.51)	(6.90)
Sample size meta-analysis = 100 (dummy)	-0.0004†	-0.01	1.75	0.0003†	-0.01†	1.96	-0.001†	-0.05	0.24	0.0002†	-0.05	1.44
	(-0.25)	(-2.93)	(21.0)	(0.02)	(-0.74)	(7.12)	(-0.24)	(-3.79)	(2.36)	(0.06)	(-6.65)	(10.2)
Sample size meta-analysis = 200 (dummy)	-0.001†	-0.01	2.29	0.001†	-0.004†	--	0.001†	-0.05	0.62	-0.002†	-0.06	2.09
	(-0.91)	(-2.71)	(22.0)	(0.03)	(-0.39)		(0.21)	(-4.17)	(5.89)	(-0.47)	(-7.28)	(10.9)
Ln average sample size in primary studies	-0.02	-0.02	-0.10	-0.01	-0.01	-0.12†	-0.03	-0.01†	-0.41	-0.01	-0.01	-0.04†
	(-9.74)	(-8.82)	(-2.52)	(-8.47)	(-3.83)	(-1.70)	(-5.70)	(-1.88)	(-7.30)	(-3.73)	(-2.55)	(-1.41)
Point-elasticity (%)[c]	-0.12	-0.16	-1.52	-0.60	-0.29	-2.26	0.20	0.15	-0.12†	0.16	0.13	-4.07
	(-73.3)	(-40.0)	(-18.1)	(-20.5)	(-19.5)	(-4.99)	(29.0)	(10.1)	(-0.81)	(35.8)	(15.9)	(-19.1)
Omitted variables (%)	0.97	1.02	6.33	1.02	0.57	2.62	0.20	0.18	1.36	0.17	0.18	-1.07
	(71.9)	(43.1)	(16.8)	(147.2)	(13.3)	(14.4)	(6.99)	(5.80)	(7.96)	(12.1)	(9.68)	(-6.77)
Log-normal distribution (%)	-0.05	-0.06	-1.39	-0.08	-0.03	-1.73	0.09	0.08	-0.39	0.08	0.07	-1.53
	(-10.5)	(-10.8)	(-18.8)	(-18.4)	(-8.16)	(-20.3)	(8.84)	(5.15)	(-3.95)	(15.4)	(9.48)	(-22.0)
Disturbance variance is 0.5 (%)	0.02	0.02	0.47	0.03	0.01	0.82	-0.07	-0.06	-0.22†	-0.02	-0.04	0.62
	(3.48)	(3.22)	(4.79)	(10.5)	(2.19)	(7.56)	(-8.77)	(-4.63)	(-1.68)	(-4.68)	(-6.28)	(10.6)
Disturbance variance is 2.0 (%)	-0.03	-0.03	-1.04	-0.05	-0.02	-1.05	0.01†	0.05	-0.47	0.05	0.05	-0.90
	(-8.36)	(-5.60)	(-12.2)	(-14.2)	(-5.29)	(-9.21)	(1.58)	(4.91)	(-2.97)	(8.46)	(6.15)	(-15.0)
N	20,000	20,000	20,000	60,000	60,000	60,000	20,000	20,000	20,000	60,000	60,000	60,000
R^2-adjusted	0.90	0.84	--	0.90	0.54	--	0.18	0.09	--	0.13	0.05	--
Log-likelihood	19771	13147	-4475	44834	20733	-2163	5445	810	-1177	14924	-11667	-11706
Restricted Log-likelihood	-3518	-5204	-10432	-25634	-2076	-3847	3469	-191	-1335	10771	-13338	-23732
F-test (Chi-squared for SIZE and POWER)	20582	11696	11915	44834	20733	3367	485	234	317	810	312	24052

[a] The results for BIAS and MSE are presented with *t*-values (in parentheses) based on clustered Huber-White corrected standard errors. Clusters are determined on the basis of different values of the population effect size and the sample size of the meta-analysis, as well as on deciles of the proportion of point-elasticities included in the meta-sample. The results for SIG are binary probit estimates, using the same corrective device for robustness. A dagger indicates that the variable is not significantly different from zero at the 5 percent level.

[b] The effects of '$\beta_0 = 0.5$', 'Sample size meta-analysis = 25,' 'Sample size primary study is 50,' and 'Disturbance variance is 1.0' are subsumed in the constant term.

[c] (%) indicates the variable is operationalised as the percentage of primary studies included in the meta-sample for which the description holds.

References

Bateman IJ, AP Jones, 2003, Contrasting Conventional with Multi-Level Modeling Approaches to Meta-Analysis: Expectation Consistency in U.K. Woodland Recreation Values, *Land Economics* **79**, 235-258.

Bijmolt THA, RGM Pieters, 2001, Meta-Analysis in Marketing when Studies Contain Multiple Measurements, *Marketing Letters* **12**, 157-169.

Card D, AB Krueger, 1995, Time-Series Minimum-Wage Studies: A Meta-Analysis, *American Economic Review* **85**, 238-243.

Dalhuisen JM, RJGM Florax, HLF de Groot, P Nijkamp, 2003, Price and Income Elasticities of Residential Water Demand: A Meta-Analysis, *Land Economics* **79**, 292-308.

Florax RJGM, 2002, Accounting for Dependence Among Study Results in Meta-Analysis: Methodology and Applications to the Valuation and Use of Natural Resources, Serie Research Memoranda, RM 2002-5, Free University, Amsterdam.

Goldfarb RS, 1995, The Economist-as-Audience Needs a Methodology of Plausible Inference, *Journal of Economic Methodology* **2**, 201-222.

Görg H, E Strobl, 2001, Multinational Companies and Productivity Spillovers: A Meta-Analysis, *Economic Journal* **111**, 723-740.

Greene WH, 2000, *Econometric Analysis*, Fourth Edition, Prentice Hall International, Inc., Upper Saddle River, New Jersey.

Hunter JE, FL Schmidt, 1990, *Methods of Meta-Analysis; Correcting Error and Bias in Research Findings*, Sage Publications, London.

Keef SP, LA Roberts, 2004, The Meta-Analysis of Partial Effect Sizes, *British Journal of Mathematical and Statistical Psychology* **57**, 97-129.

Oswald FL, JW Johnson, 1998, On the Robustness, Bias, and Stability of Statistics from Meta-Analysis of Correlation Coefficients: Some Initial Monte Carlo Findings, *Journal of Applied Psychology* **83**, 164-178.

Rosenthal R, MR DiMatteo, 2001, Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews, *Annual Review of Psychology* **52**, 59-82.

Smith VK, JC Huang, 1995, Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models, *Journal of Political Economy* **103**, 209-227.

Smith VK, SK Pattanayak, 2002, Is Meta-Analysis a Noah's Ark for Non-Market Valuation?, *Environmental and Resource Economics* **22**, 271-296.

Stanley TD, 2001, Wheat From Chaff: Meta-Analysis As Quantitative Literature Review, *Journal of Economic Perspectives* **15**, 131-150.

Sutton AJ, KR Abrams, TA Sheldon, F Song, 2000, *Methods for Meta-Analysis in Medical Research*, John Wiley and Sons, New York.

Weichselbaumer D, R Winter-Ebmer, 2003, A Meta-Analysis of the International Gender Wage Gap, CEPR discussion paper, no. 4127, London.

Williams RL, 2000, A Note on Robust Variance Estimation for Cluster-Correlated Data, *Biometrics* **56**, 645-646.

Wooldridge JM, 2002, *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge, Massachusetts.